

University of Groningen

## Consistent and asymptotically normal PLS estimators for linear structural equations

Dijkstra, Theo K.; Henseler, Jörg

*Published in:*  
Computational Statistics & Data Analysis

*DOI:*  
[10.1016/j.csda.2014.07.008](https://doi.org/10.1016/j.csda.2014.07.008)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2015

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis*, 81(1), 10-23.  
<https://doi.org/10.1016/j.csda.2014.07.008>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



# Consistent and asymptotically normal PLS estimators for linear structural equations

Theo K. Dijkstra<sup>a,\*</sup>, Jörg Henseler<sup>b</sup>

<sup>a</sup> University of Groningen, NL, Department of Economics and Econometrics, The Netherlands

<sup>b</sup> University of Twente, NL, Department of Design, Production and Management, The Netherlands

## HIGHLIGHTS

- Consistent PLS estimates path coefficients and indicator loadings consistently.
- Consistent PLS can estimate parameters of nonrecursive structural equation models.
- A family of goodness-of-fit measures makes PLS suitable for confirmatory research.
- Consistent PLS performs comparably to covariance-based structural equation modeling.

## ARTICLE INFO

### Article history:

Received 22 December 2013

Received in revised form 16 July 2014

Accepted 17 July 2014

Available online 24 July 2014

### Keywords:

Partial least squares

Structural equation modeling

Consistency

Recursiveness

Goodness-of-fit

## ABSTRACT

A vital extension to partial least squares (PLS) path modeling is introduced: consistency. While maintaining all the strengths of PLS, the consistent version provides two key improvements. Path coefficients, parameters of simultaneous equations, construct correlations, and indicator loadings are estimated consistently. The global goodness-of-fit of the structural model can also now be assessed, which makes PLS suitable for confirmatory research. A Monte Carlo simulation illustrates the new approach and compares it with covariance-based structural equation modeling.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## 1. Introduction

Structural equation modeling (SEM) has become *the* tool of the trade in survey-based research. Researchers embrace its abilities, such as modeling latent variables and correcting for measurement error, while simultaneously estimating parameters of entire theories. Two families of structural equation modeling techniques prevail (Chin, 1998; Reinartz et al., 2009): covariance-based SEM and variance-based SEM. The latter appears to be increasingly popular, as seen in recent methodological advances (cf. Bry et al., 2012; Hwang et al., 2010; Lu et al., 2011; Tenenhaus and Tenenhaus, 2011) and frequent application (Ringle et al., 2012; Hair et al., 2012). Researchers appreciate the advantages of variance-based SEM, such as the lack of convergence problems and factor indeterminacy (Henseler, 2010), relatively mild distributional assumptions (Reinartz et al., 2009), and the possibility of estimating models having more variables or parameters than observations. Variance-based SEM includes different techniques, such as regression based on sum scores or principal components (Tenenhaus, 2008), partial least squares path modeling (PLS, see Wold, 1982; Tenenhaus et al., 2005), and

\* Correspondence to: University of Groningen, Faculty of Economics and Business, Nettelbosje 2, 9747 AE Groningen, The Netherlands. Tel.: +31 50 363 3793; fax: +31 50 363 7337.

E-mail addresses: [t.k.dijkstra@rug.nl](mailto:t.k.dijkstra@rug.nl) (T.K. Dijkstra), [joerg@henseler.com](mailto:joerg@henseler.com) (J. Henseler).

<http://dx.doi.org/10.1016/j.csda.2014.07.008>

0167-9473/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

generalized structured component analysis (Hwang et al., 2010; Henseler, 2012). All variance-based SEM techniques have the characteristic of approximating latent variables using linear composites of observed variables.

Among variance-based SEM techniques, PLS path modeling is regarded as the “most fully developed and general system” (McDonald, 1996, p. 240). Various extensions and advances of PLS path modeling have been developed and discussed, such as multigroup analysis (Chin and Dibbern, 2010; Sarstedt et al., 2011), testing moderating effects (Chin et al., 2003; Henseler and Chin, 2010), assessing common method bias (Liang et al., 2007; Chin et al., forthcoming), testing measurement invariance (Hsieh et al., 2008; Money et al., 2012), modeling nonlinear relationships (Henseler et al., 2012), and analyzing hierarchical component models (Ringle et al., 2012; Wetzels et al., 2009). The large number of extensions and advances are in line with the widespread dissemination of PLS path modeling across business disciplines such as management information systems (cf. Ringle et al., 2012), marketing (cf. Hair et al., 2012; Henseler et al., 2009), strategy (cf. Hulland, 1999), or operations management (cf. Peng and Lai, 2012).

Yet, the use of PLS as an estimator for structural equation models is not without disadvantages. First, PLS estimates – particularly path coefficients and loadings – are only consistent at large (Wold, 1982). Consequently, “(p)arameter estimates for paths between observed variables and latent variable proxies are biased upward in PLS (away from zero), while parameter estimates for paths between proxies are attenuated” (Gefen et al., 2011, p. vi). Second, PLS does not provide overall goodness-of-fit measures. This means it is impossible to test or compare theories, as is done with covariance-based SEM, is not possible (Fornell and Bookstein, 1982; Henseler and Sarstedt, 2013). Of these two deficiencies of PLS, the lack of consistency is probably more serious, because of its adverse consequences for substantial research findings. If PLS underestimates the true parameter, Type II errors are likely. If PLS overestimates the true parameter, the Type I error is inflated. Finally, the lack of consistency entails that there is no guarantee that meta studies based on PLS estimates come closer to the true value than single studies.

Against this background, we introduce an important advancement to PLS that overcomes these deficiencies: consistent PLS. While maintaining all the strengths of PLS, consistent PLS provides several key improvements. It consistently estimates the path coefficients, construct correlations, and indicator loadings; it allows non-recursive models to be estimated; and it provides a global assessment of goodness-of-fit. The outcomes of a Monte Carlo simulation help compare the performance of PLS, consistent PLS, and covariance-based SEM. Finally, implications for future research are provided.

## 2. Consistent partial least squares

Scandinavian econometrician and statistician Herman Wold developed partial least squares in the 1960s, 1970s, and 1980s to analyze high-dimensional data that reflects unobserved entities that are interconnected in path diagrams (Wold, 1966, 1975, 1982). The inspiration were principal components and canonical variables. Linear compounds are constructed to serve as proxies or stand-ins for the latent variables, leading to straightforward estimates of structural parameters, such as path coefficients and loadings. Compound weights are generated, using a variety of alternating least squares algorithms. These are cycles of regressions that update natural subsets of weights in turn, stopping when consecutive estimates no longer change significantly. Convergence, which is the rule, is typically very fast. PLS has become a vibrant field of both applied and theoretical research; see, for example, Tenenhaus et al. (2005) and the Handbook of Partial Least Squares by Vinzi et al. (2010) for overviews.

In the spirit of principal components and canonical variables, PLS has been useful as a method to extract information from high-dimensional data. However, as a means of estimating the parameters of latent variable models, PLS has shortcoming: the relationships between linear compounds can never duplicate the relationships between the latent variables. The simple and fundamental reason is that no linear combination of the indicators of a block can ever replicate the corresponding latent variable, except when some measurement errors are zero (for an analysis of the general case see Krijnen et al., 1998). See Section 2.3 and Eq. (14) for the fundamental relationship between the correlations among the proxies, the true correlations among the latent variables, and those between each proxy and corresponding latent variable. In fact, in linear factor models, PLS tends to overestimate the absolute value of loadings and underestimate the multiple and bivariate (absolute) correlation coefficients. Dijkstra (1981, 1983, 2010) shows how to correct for this tendency. The consistent version of PLS is denoted by PLSc. Subsequent sections outline the PLSc approach and show that it gives consistent and asymptotically normal estimators (CAN estimators) for the focal parameters.

### 2.1. Weight vectors

A starting point for PLS analysis is the “basic design”, in essence a second-order factor model. A number of i.i.d. (independent and identically distributed) vectors of indicators are assumed to exist from a population with finite moments of at least order two (the precise order depends on other distributional assumptions or requirements). All indicators have zero mean and unit variance. The vector of indicators  $y$  is composed of at least two subvectors, each measuring a unique latent variable, and each subvector contains at least two components. For the  $i$ th subvector  $y_i$  we have

$$y_i = \lambda_i \cdot \eta_i + \epsilon_i \quad (1)$$

where the loading vector  $\lambda_i$  and the vector of idiosyncratic errors  $\epsilon_i$  have the same dimensions as  $y_i$ , and the unobservable latent variable  $\eta_i$  is real-valued. For convenience, the sufficient but by no means necessary assumption is made that all

components of all error vectors are mutually independent, and independent of all latent variables. The latter has zero mean and unit variance. The correlations between  $\eta_i$  and  $\eta_j$  are denoted by  $\rho_{ij}$ . They are collected in a matrix  $\Phi$ ,  $\Phi := (\rho_{ij})$ . At this stage, the nature of the relationships between the latent variables – whether linear or nonlinear – is not relevant.

A particular set of easy implications is that the covariance matrix  $\Sigma_{ii}$  of  $y_i$  can be written as

$$\Sigma_{ii} := E y_i y_i^\top = \lambda_i \lambda_i^\top + \Theta_i \quad (2)$$

where  $\Theta_i$ , the covariance matrix of the measurement errors of the  $i$ th latent variable, is diagonal with non-negative diagonal elements, and we have for the covariance between  $y_i$  and  $y_j$

$$\Sigma_{ij} := E y_i y_j^\top = \rho_{ij} \lambda_i \lambda_j^\top. \quad (3)$$

The sample counterparts of  $\Sigma_{ii}$  and  $\Sigma_{ij}$  are denoted by  $S_{ii}$  and  $S_{ij}$ , respectively. The sample data is assumed to be standardized before being analyzed. Therefore, the observed data has zero mean and unit (sample) variance. Note that the assumptions made so far entail that the sample counterparts are consistent and asymptotically normal estimators of the theoretical variance and covariance matrices.

PLS features a number of iterative *fixed-point* algorithms, of which the mode A algorithm is selected. In general, the mode A algorithm is numerically the most stable algorithm (for discussions of other PLS modes, see Lohmöller, 1989). As a rule, the algorithm converges and is usually very fast (for example, for the models analyzed in this paper, no more than five iterations are needed to obtain five significant decimals). The outcome is an estimated weight vector  $\hat{w}$  with typical subvector  $\hat{w}_i$  of the same dimensions as  $y_i$ . With these weights, *sample proxies* are defined for the latent variables:  $\hat{\eta}_i := \hat{w}_i^\top y_i$  for  $\eta_i$ , with the customary normalization of a unit sampling variance, so  $\hat{w}_i^\top S_{ii} \hat{w}_i = 1$ . In Wold's PLS approach, the  $\hat{\eta}_i$ 's replace the unobserved latent variables, and the loadings and structural parameters are estimated directly, in contrast to covariance-based SEM, which, for example, follows the opposite order. In mode A, for each  $i$ :

$$\hat{w}_i \propto \sum_{j \in C(i)} \text{sgn}_{ij} \cdot S_{ij} \hat{w}_j. \quad (4)$$

Here,  $\text{sgn}_{ij}$  is the sign of the sample correlation between  $\hat{\eta}_i$  and  $\hat{\eta}_j$  and  $C(i)$  is a set of indices of latent variables. Traditionally,  $C(i)$  contains the indices of latent variables adjacent to  $\eta_i$ , in other words, those that appear on the *other* side of the structural or path equations in which  $\eta_i$  appears. This setup is not always a good idea, particularly when the correlations between the indicators of  $\eta_i$  and those of its adjacent variables are weak. In general, the use of *all*  $j \neq i$  is suggested in Eq. (4). Clearly,  $\hat{w}_i$  is obtained by a regression of the indicators  $y_i$  on the sign-weighted sum of the selected proxies:  $\sum_{j \in C(i)} \text{sgn}_{ij} \cdot \hat{\eta}_j$ . Other versions exist (for example, with correlation weights); this version, which is the original, is one of the simplest (see Wold, 1982). There is little motivation in the PLS literature for the coefficients of  $S_{ij} \hat{w}_j$ . The particular choice can be shown to be irrelevant for the probability limits of the estimators. The algorithm takes an arbitrary starting vector and then basically follows the sequence of regressions for each  $i$ , each time inserting updates when available (or after each full round; the precise implementation is not important).

Dijkstra (1981, 2010) showed that the PLS modes converge with a probability that tends to one when the sample size tends to infinity for essentially arbitrary starting vectors. The weight vectors that satisfy the fixed-point equations are locally continuously differentiable functions of the sample covariance matrix of  $y$ . Therefore, they and other estimators that depend smoothly on the weight vectors and  $S$  are jointly asymptotically normal.

## 2.2. Factor loadings

Let us denote the probability limit of  $\hat{w}_i$ ,  $\text{plim } \hat{w}_i$ , by  $\bar{w}_i$ . We can get it from Eq. (4) by substituting  $\Sigma$  for  $S$ . So

$$\bar{w}_i \propto \sum_{j \in C(i)} \text{sgn}_{ij} \cdot \Sigma_{ij} \bar{w}_j = \sum_{j \in C(i)} \text{sgn}_{ij} \cdot \rho_{ij} \cdot \lambda_i \cdot \lambda_j^\top \bar{w}_j = \lambda_i \sum_{j \in C(i)} \text{sgn}_{ij} \cdot \rho_{ij} \cdot \lambda_j^\top \bar{w}_j \quad (5)$$

where the last sum is just a real number, and we conclude that  $\bar{w}_i$  is proportional to  $\lambda_i$ . Because of the normalization (unit variance) the proportionality constant has to be such that  $\bar{w}_i^\top \Sigma_{ii} \bar{w}_i = 1$ . This entails that

$$\bar{w}_i = \frac{\lambda_i}{\sqrt{\lambda_i^\top \Sigma_{ii} \lambda_i}}. \quad (6)$$

In PLS, the loadings are estimated by a regression of the indicators  $y_i$  on their direct sample proxy  $\hat{\eta}_i$ . However, because doing so in general removes the proportionality, *this tradition is not followed (for mode A)*.

As in Dijkstra (1981, 2010), the following estimator for  $\lambda_i$  is proposed:

$$\hat{\lambda}_i := \hat{c}_i \cdot \hat{w}_i, \quad (7)$$

where the scalar  $\hat{c}_i$  is such that the off-diagonal elements of  $S_{ii}$  are reproduced as best as possible in a least squares sense. Therefore, the Euclidean distance is minimized between:

$$\left[ S_{ii} - \text{diag}(S_{ii}) \right] \quad \text{and} \quad \left[ (c_i \cdot \hat{w}_i) (c_i \cdot \hat{w}_i)^\top - \text{diag}((c_i \cdot \hat{w}_i) (c_i \cdot \hat{w}_i)^\top) \right]$$

as a function of  $c_i$  and the following is obtained:

$$\widehat{c}_i := \sqrt{\frac{\widehat{w}_i^\top (S_{ii} - \text{diag}(S_{ii})) \widehat{w}_i}{\widehat{w}_i^\top (\widehat{w}_i \widehat{w}_i^\top - \text{diag}(\widehat{w}_i \widehat{w}_i^\top)) \widehat{w}_i}}. \quad (8)$$

More explicitly,

$$\widehat{c}_i = \left[ \frac{\sum_{a \neq b} \widehat{w}_{a,i} \widehat{w}_{b,i} S_{ii,ab}}{\sum_{a \neq b} \widehat{w}_{a,i}^2 \widehat{w}_{b,i}^2} \right]^{\frac{1}{2}}, \quad (9)$$

where  $\widehat{w}_{a,i}$  is element ‘a’ of the weight vector  $\widehat{w}_i$ ,  $\widehat{w}_{b,i}$  is defined similarly, and  $S_{ii,ab}$  is the sample covariance between the indicators ‘a’ and ‘b’ of the  $i$ th block. When pairs of errors of the same block are suspected to be correlated, one can delete the corresponding terms in both numerator and denominator.<sup>1</sup>

In sufficiently large samples,  $\widehat{c}_i$  will be well-defined, real, and positive. (In all samples in this paper and those in another study,  $\widehat{c}_i$  attained proper values.) Its calculation does not require additional numerical optimization. Verifying that the correction does its job is straightforward by replacing  $S_{ii}$  by  $\Sigma_{ii}$  and  $\widehat{w}_i$  by  $\overline{w}_i$ : the matrix in the denominator equals the matrix in the numerator, apart from a factor  $(\lambda_i^\top \Sigma_{ii} \lambda_i)^{-1}$ ; therefore:

$$\overline{c}_i := \text{plim} \widehat{c}_i = \sqrt{\lambda_i^\top \Sigma_{ii} \lambda_i}. \quad (10)$$

Now, in particular

$$\text{plim} \widehat{\lambda}_i = \text{plim} (\widehat{c}_i \cdot \widehat{w}_i) = \overline{c}_i \cdot \overline{w}_i = \lambda_i. \quad (11)$$

### 2.3. Correlations between latent variables

Defining a population proxy  $\overline{\eta}_i$  by  $\overline{\eta}_i := \overline{w}_i^\top y_i$  is useful. Clearly, the squared correlation between a population proxy and its corresponding latent variable is

$$R^2(\eta_i, \overline{\eta}_i) = (\overline{w}_i^\top \lambda_i)^2, \quad (12)$$

which equals

$$\frac{(\lambda_i^\top \lambda_i)^2}{\lambda_i^\top \Sigma_{ii} \lambda_i} = \frac{(\lambda_i^\top \lambda_i)^2}{(\lambda_i^\top \lambda_i)^2 + \lambda_i^\top \Theta_i \lambda_i}. \quad (13)$$

With a “large” number of “high quality” indicators, this correlation can be close to one (“consistency at large” in PLS parlance). A trivially deduced but important algebraic relationship is

$$R^2(\overline{\eta}_i, \overline{\eta}_j) = (\overline{w}_i^\top \Sigma_{ij} \overline{w}_j)^2 = \rho_{ij}^2 \cdot R^2(\eta_i, \overline{\eta}_i) \cdot R^2(\eta_j, \overline{\eta}_j). \quad (14)$$

Because proxies can never replicate the latent variables exactly, barring special situations where measurement errors are identically zero, entails that their correlation matrix can never equal the one that describes the relationships between the latent variables. The probability limits of the bivariate squared correlations are too small, as are those of the multiple correlations (see Dijkstra, 2010). In addition, regression coefficients and coefficients of simultaneous equations are misrepresented. See Section 3.3 for a numerical example.

Now note that

$$R^2(\eta_i, \overline{\eta}_i) = (\overline{w}_i^\top \lambda_i)^2 = (\overline{w}_i^\top \cdot (\overline{w}_i \cdot \overline{c}_i))^2 = (\overline{w}_i^\top \overline{w}_i)^2 \cdot \overline{c}_i^2 \quad (15)$$

enabling an estimation of the (squared) quality of the proxies consistently by

$$R^2(\widehat{\eta}_i, \widehat{\eta}_i) := (\widehat{w}_i^\top \widehat{w}_i)^2 \cdot \widehat{c}_i^2. \quad (16)$$

Also, with

$$R^2(\widehat{\eta}_i, \widehat{\eta}_j) := (\widehat{w}_i^\top S_{ij} \widehat{w}_j)^2, \quad (17)$$

<sup>1</sup> For extensions and alternative approaches, see Dijkstra (2013).

see Eq. (14); the correlations between the latent variables can be consistently estimated using

$$\hat{\rho}_{ij}^2 := \frac{R^2(\hat{\eta}_i, \hat{\eta}_j)}{R^2(\eta_i, \hat{\eta}_i) \cdot R^2(\eta_j, \hat{\eta}_j)}. \quad (18)$$

Therefore,  $E\eta_i\eta_j$  is estimated using the sample covariance between the proxies  $\hat{\eta}_i$  and  $\hat{\eta}_j$ , each divided by its estimated quality. Finally, let  $\hat{\Phi} := (\hat{\rho}_{ij})$ .

Note that standard PLS software for mode A produces all the necessary ingredients for a consistent estimation; all that is required is a simple rescaling of the weight vectors.<sup>2</sup>

#### 2.4. Simultaneous equation systems

It is important to note that when the latent variables are mutually related through linear equations, whether recursively or with feedback patterns, CAN estimates of their coefficients are also obtainable. This can happen if they are identifiable from the second-order moment matrix of the latent variables through smooth (locally continuously differentiable) mappings. In principle, partially identifiable structures can also be handled using Bekker and Dijkstra (1990) and Bekker et al. (1994).

The method of choice in this study is the old econometric workhorse 2SLS (two-stage least squares). 2SLS estimates each equation separately and is a limited-information technique. It is probably the simplest estimation method around, and does not – despite appearances – require additional iterations. Boardman et al. (1981) employed an iterative version of 2SLS, Wold's fix-point method, using the original PLS input, not corrected for inconsistency.

To be complete, the 2SLS method is specified here. To this end, consider the linear structural equations

$$\eta_{p+1:p+q} = B \cdot \eta_{p+1:p+q} + \Gamma \cdot \eta_{1:p} + \zeta, \quad (19)$$

where  $\eta$  is partitioned into a vector of  $p$  components,  $\eta_{1:p}$ , the exogenous latent variables, and a vector of  $q$  components,  $\eta_{p+1:p+q}$ , the endogenous latent variables. The residual vector  $\zeta$  has zero mean and is uncorrelated with (or independent of)  $\eta_{1:p}$ . The  $q \times q$  matrix  $B$  captures the feedback or reciprocal relationships between the endogenous variables, and is such that the inverse of  $I - B$  exists. The latter assumption enables one to write, with  $\Pi := (I - B)^{-1} \Gamma$

$$\eta_{p+1:p+q} = \Pi \cdot \eta_{1:p} + (I - B)^{-1} \zeta, \quad (20)$$

which is a set of  $q$  regression equations. *Identifiability* is assumed, which means that  $B$  and  $\Gamma$  satisfy zero constraints that allow the unambiguous recovery of the values of their free parameters from the knowledge of  $\Pi$ . As is well-known, this is equivalent to the specification of the ranks of certain sub-matrices of  $\Pi$  and to the invertibility of a certain matrix, as specified below. The observation that led to 2SLS is that

$$\eta_{p+1:p+q} = B \cdot (\Pi \cdot \eta_{1:p}) + \Gamma \cdot \eta_{1:p} + (I - B)^{-1} \zeta \quad (21)$$

(with  $\Gamma = (I - B) \cdot \Pi$ ). Therefore, the free elements in a row of  $B$  and  $\Gamma$  are regression coefficients. They can be obtained through a regression of the corresponding endogenous variable on the predicted values of the endogenous variables on the right-hand side of the equation (the relevant elements of  $\Pi \cdot \eta_{1:p}$ ), and the exogenous variables of the equation, which are the relevant elements of  $\eta_{1:p}$ .

The solutions for the  $i$ th row are spelled out. Let  $I_i$  select the free parameters in the  $i$ th row of  $B$  (therefore,  $I_i$  is a vector containing the positions in the  $i$ th row of  $B$  corresponding to free parameters), and let  $J_i$  be defined similarly for the  $i$ th row of  $\Gamma$ . Therefore, the column vector of free parameters in the  $i$ th row of  $B$ , denoted by  $\beta_i$ , equals  $B(i, I_i)^\top$ , and for the free parameters in the  $i$ th row of  $\Gamma$ , we define  $\gamma_i := \Gamma(i, J_i)^\top$ . Then

$$\begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} = \begin{bmatrix} \text{cov}(\Pi(I_i, 1:p) \eta_{1:p}) & E(\Pi(I_i, 1:p) \eta_{1:p} \cdot \eta_{J_i}^\top) \\ \text{cov}(\eta_{J_i}) \end{bmatrix}^{-1} \begin{bmatrix} E(\Pi(I_i, 1:p) \eta_{1:p} \cdot \eta_{p+i}) \\ E(\eta_{J_i} \cdot \eta_{p+i}) \end{bmatrix}. \quad (22)$$

Because of identifiability, the matrix inverted is invertible. Using straightforward algebra leads to the following:

$$\begin{aligned} \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} &= \begin{bmatrix} \Phi(p + I_i, 1:p) \Phi(1:p, 1:p)^{-1} \Phi(1:p, p + I_i) & \Phi(p + I_i, J_i) \\ \Phi(J_i, p + I_i) & \Phi(J_i, J_i) \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} \Phi(p + I_i, 1:p) \Phi(1:p, 1:p)^{-1} \Phi(1:p, p + i) \\ \Phi(J_i, p + i) \end{bmatrix}. \end{aligned} \quad (23)$$

<sup>2</sup> An alternative approach, even simpler than PLS, would be to use fixed weights without iterations. See Dijkstra (2013) and Dijkstra and Schermelleh-Engel (2013), Section 2.4. The estimators of structural parameters and loadings, etc. will depend on the weights chosen. This is unlike PLS, which neutralizes the initial choice by iterating until a fixed point appears. Since PLS is so fast, the numerical gains may be modest, but the statistical merits of the fixed weights certainly need to be investigated. They are currently examined in the context of non-linear factor models.

$\Phi$  is the covariance (correlation) matrix of  $\eta = [\eta_{1:p}; \eta_{p+1:p+q}]$ , which is taken to be invertible (no redundant latent variables).

Eq. (23) clarifies how to obtain CAN estimators for the parameters of the structural equations: simply replace  $\Phi$  by its CAN estimator derived in the previous section. The ensuing vector, with components  $\hat{\beta}_i$  and  $\hat{\gamma}_i$ , is a smooth transformation (in the neighborhood of the true values) of  $\hat{\Phi}$ . Evidently, straightforward estimators (direct sample counterparts) for  $\Pi$  and for the covariance matrices of the residuals (for the structural and reduced form) share the asymptotic properties. In fact, all parameter estimators derived so far, plus the implied estimator for  $\Sigma$ , are consistent and asymptotically jointly normal. They will not be asymptotically most efficient when the sample comes from a distribution like the Gaussian or an elliptical distribution: neither the weights, loadings, and correlations, nor the structural form coefficients are determined by taking all information optimally into account. There is also an advantage, although this study does not elaborate on it: full-information methods are potentially vulnerable to misspecification anywhere in the system; the approach outlined here can be expected to be more robust.

### 2.5. Standard errors and tests-of-fit

Because of the speed of PLS, simulating the distribution of the estimators on the basis of the empirical distribution of the sample is quite feasible. Correction of bias, if any, and estimation of standard errors and confidence intervals are, in principle, well within reach. Some simulation results are reported below. Alternatively, one may use the delta method with the Jacobian matrix calculated numerically. One can then obtain estimates of the standard errors based on Gaussian or distribution-free asymptotic theory (through higher-order moments or the bootstrap).

As the implied  $\hat{\Sigma}$  based on direct substitution of the corrected PLS estimators is consistent and asymptotically normal, one may consider the use of overall tests as in the covariance-based SEM literature by defining a proper distance such as the trace of the square of the residual matrix  $S - \hat{\Sigma}$ . When scaled by the number of observations, the distance is distributed asymptotically as a non-negative linear combination of independent  $\chi^2(1)$ -variables. The coefficients are eigenvalues of a certain matrix that depends on the true parameters. One could replace them with appropriate estimates and employ a suitable approximation for the probability value. Alternatively, and more conveniently, the probability value can be estimated using the bootstrap. This requires a pre-multiplication of the observation vectors by  $\hat{\Sigma}^{-\frac{1}{2}} S^{-\frac{1}{2}}$ , meaning that the covariance matrix of their empirical distribution satisfies the assumed ( $H_0$ ) structure (see, for example, Yuan and Hayashi, 2003, for a general discussion and elaboration in the context of covariance analysis). We report some simulation results below.

## 3. Monte Carlo experiment

To assess the quality of the estimators provided by consistent PLS, a computational experiment was conducted. In particular, the performance of consistent PLS is compared with covariance-based SEM, using a Monte Carlo simulation.

### 3.1. Setup

The first illustrative test asks for a non-trivial model that is not too large. The experiment should be challenging, meaning that the sample size should be modest and the number of indicators small. It is investigated how accurate the limited information method PLS performs when pitted against the most efficient alternative—full information maximum likelihood (FIML). The effect of non-normality will also be investigated.

The model chosen is Summers' (1965) classical model, which is used in many econometrics studies to directly observe endogenous and exogenous variables. Here, a latent vector is composed of six components,  $\eta_{1:6}$ , linked to each other through the following equations:

$$\eta_{5:6} = B\eta_{5:6} + \Gamma\eta_{1:4} + \zeta, \quad (24)$$

where  $\zeta$  is a two-dimensional residual vector independent of  $\eta_{1:4}$ , and the coefficient matrices take the following forms:

$$B := \begin{bmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{bmatrix} \quad (25)$$

and

$$\Gamma := \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 \\ 0 & 0 & \gamma_{23} & \gamma_{24} \end{bmatrix}. \quad (26)$$

Spelled out:

$$\eta_5 = \beta_{12}\eta_6 + \gamma_{11}\eta_1 + \gamma_{12}\eta_2 + \zeta_1 \quad (27)$$

$$\eta_6 = \beta_{21}\eta_5 + \gamma_{23}\eta_3 + \gamma_{24}\eta_4 + \zeta_2. \quad (28)$$



The endogenous variables influence one another reciprocally (for a causal interpretation of these equations, see Pearl, 2009, who builds on Haavelmo, 1944). The structural form equations are not regressions, but the reduced form equations are, with

$$\Pi = \frac{1}{1 - \beta_{12}\beta_{21}} \begin{bmatrix} \gamma_{11} & \gamma_{12} & \beta_{12}\gamma_{23} & \beta_{12}\gamma_{24} \\ \beta_{21}\gamma_{11} & \beta_{21}\gamma_{12} & \gamma_{23} & \gamma_{24} \end{bmatrix}. \quad (29)$$

Of course,  $1 - \beta_{12}\beta_{21} \neq 0$  is required. All structural form coefficients can be recovered unambiguously from  $\Pi$  if and only if each of the submatrices  $\Pi(1 : 2, 1 : 2)$  and  $\Pi(1 : 2, 3 : 4)$  has rank one. Therefore,  $\gamma_{11}$  and  $\gamma_{12}$  cannot both be zero, nor can  $\gamma_{23}$  and  $\gamma_{24}$ . The following coefficients were chosen:

$$B = \begin{bmatrix} 0 & 0.25 \\ 0.50 & 0 \end{bmatrix} \quad (30)$$

and

$$\Gamma = \begin{bmatrix} -0.30 & 0.50 & 0 & 0 \\ 0 & 0 & 0.50 & 0.25 \end{bmatrix}. \quad (31)$$

All covariances (correlations) between the exogenous latent variables  $\eta_{1:4}$  are equal to 0.5 and the correlation between  $\eta_5$  and  $\eta_6$  is  $\sqrt{0.5}$ . All numbers are chosen arbitrarily. The regression matrix equals

$$\Pi = \begin{bmatrix} -0.3429 & 0.5714 & 0.1429 & 0.0714 \\ -0.1714 & 0.2857 & 0.5714 & 0.2857 \end{bmatrix}. \quad (32)$$

The correlation matrix between the latent variables can be verified as

$$\Phi = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.0500 & 0.4000 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5071 & 0.6286 \\ 0.5 & 0.5 & 1 & 0.5 & 0.2929 & 0.7714 \\ 0.5 & 0.5 & 0.5 & 1 & 0.2571 & 0.6286 \\ 0.0500 & 0.5071 & 0.2929 & 0.2571 & 1 & \sqrt{0.5} \\ 0.4000 & 0.6286 & 0.7714 & 0.6286 & \sqrt{0.5} & 1 \end{bmatrix}. \quad (33)$$

Note that the correlation between  $\eta_5$  and  $\eta_1$  is particularly weak (0.05), which may create a challenge for PLSc.

One can verify that the R-squared for the first reduced form equation is a mere 0.3329, whereas the corresponding value for the second equation is a healthy 0.7314. The implied value for the covariance matrix of  $\zeta$  is

$$\Sigma_{\zeta\zeta} = \begin{bmatrix} 0.5189 & -0.0295 \\ -0.0295 & 0.1054 \end{bmatrix}. \quad (34)$$

As for the  $\lambda$ 's, the main experiment takes just three indicators per latent variable, and all components of all loading vectors equal 0.70 (making their squared correlation with their corresponding latent variable less than one half).

### 3.2. Distributions

The model contains 31 free parameters—two from  $B$ , four from  $\Gamma$ , six from the correlation matrix of  $\eta_{1:4}$ , one correlation between  $\zeta_1$  and  $\zeta_2$ , and 18 from the loadings. Using the normalizations, all other parameters can be derived from the ones referred to. Therefore, a sample size  $n$  of 300 with about 10 observations per parameter seems modest. For some experiments,  $n = 600$  and  $n = 1200$ . The leading distribution is the multivariate normal distribution. With the parameters specified as above, we can determine the covariance matrix  $\Sigma$  of size  $18 \times 18$  and generate  $n$  random drawings by  $\text{randn}(n, 18) \cdot \Sigma^{\frac{1}{2}}$ . The observation vectors are standardized and fed to PLSc, 2SLS is employed, and outcomes noted. This process is repeated 10,000 times.

It is customary to study the effects of non-normality by the Fleishman–Vale–Maurelli procedure (see Fleishman, 1978; Vale and Maurelli, 1983). In this approach, the standard normal latent variables are replaced by well-chosen linear combinations of powers of standard normal variables whose correlations are such that the new latent variables have the same correlations as the original latent variables. “Well-chosen” means that specified requirements concerning the non-normal skewness and (excess) kurtosis are satisfied. If the transformations as suggested maintain the independence between latent variables and idiosyncratic errors, the asymptotic robustness of normal-theory statistics may apply and lead one to believe incorrectly that normality is not an issue (see Hu et al., 1992). We follow the latter authors in simply rescaling the vector of indicators by multiplying each component by the same independent random factor. Hu et al. (1992) chose  $\sqrt{3} \cdot (\chi^2(5))^{-\frac{1}{2}}$ , whose squared value has expectation one. This approach deliberately destroys the independence between the latent and idiosyncratic variables, but leaves  $\Sigma$  and the linear relationships (as well as the symmetry) undisturbed. The kurtosis of the indicators increases by six. The same effects can be obtained by multiplying by a standard normal variable  $Z$ , which appears to yield representative samples for smaller sizes; therefore, this approach is used. In addition, multiplication is employed by a positive scale factor  $\sqrt{\text{abs}(Z) \cdot \sqrt{\frac{\pi}{2}}}$  whose squared value also has expectation one. This multiplication increases the kurtosis by a lesser amount:  $\frac{3\pi}{2} - 3 = 1.7124$ .



**Table 1**  
Unrestricted latent variable correlations obtained through PLSc.

Latent variable correlation	Population value	Estimates		Bootstrap std	
		Mean	Std	Mean	Std
$\rho_{12}$	0.5000	0.4993	0.0639	0.0639	0.0048
$\rho_{13}$	0.5000	0.4990	0.0639	0.0640	0.0054
$\rho_{14}$	0.5000	0.4997	0.0640	0.0641	0.0054
$\rho_{15}$	0.0500	0.0535	0.0798	0.0792	0.0051
$\rho_{16}$	0.4000	0.4020	0.0691	0.0682	0.0050
$\rho_{23}$	0.5000	0.5000	0.0654	0.0639	0.0051
$\rho_{24}$	0.5000	0.5006	0.0649	0.0640	0.0053
$\rho_{25}$	0.5071	0.5060	0.0645	0.0633	0.0052
$\rho_{26}$	0.6286	0.6286	0.0581	0.0567	0.0053
$\rho_{34}$	0.5000	0.5004	0.0650	0.0642	0.0052
$\rho_{35}$	0.2929	0.2951	0.0724	0.0715	0.0050
$\rho_{36}$	0.7714	0.7709	0.0480	0.0474	0.0052
$\rho_{45}$	0.2571	0.2590	0.0740	0.0726	0.0049
$\rho_{46}$	0.6286	0.6293	0.0581	0.0572	0.0054
$\rho_{56}$	0.7071	0.7049	0.0530	0.0520	0.0054

### 3.3. Traditional PLS

Although a comparison with the traditional PLS method is not the main goal of this paper, specifying the probability limits for the estimators of PLS mode A seems appropriate for the main parameters. Direct calculation using the true covariance matrix  $\Sigma$  and 2SLS for the structural parameters yields the following probability limits:

- for the loadings (regression of indicators on proxy): 0.8124 instead of 0.7
- for the correlations between the exogenous latent variables: 0.3712 instead of 0.5
- for the correlation between the endogenous latent variables: 0.5250 instead of 0.7071 ( $=\sqrt{0.5}$ )
- for  $B$ :  $\begin{bmatrix} 0 & 0.2927 \\ 0.5938 & 0 \end{bmatrix}$  instead of  $\begin{bmatrix} 0 & 0.25 \\ 0.50 & 0 \end{bmatrix}$
- for  $\Gamma$ :  $\begin{bmatrix} -0.1611 & 0.2997 & 0 & 0 \\ 0 & 0.3624 & 0.2188 & 0 \end{bmatrix}$  instead of  $\begin{bmatrix} -0.30 & 0.50 & 0 & 0 \\ 0 & 0 & 0.50 & 0.25 \end{bmatrix}$
- for  $\Pi$ :  $\begin{bmatrix} -0.1949 & 0.3628 & 0.1284 & 0.0775 \\ -0.1158 & 0.2154 & 0.4386 & 0.2648 \end{bmatrix}$  instead of  $\begin{bmatrix} -0.3429 & 0.5714 & 0.1429 & 0.0714 \\ -0.1714 & 0.2857 & 0.5714 & 0.2857 \end{bmatrix}$ .
- The implied squared correlations for the two reduced form equations are 0.1726 and 0.4421, to be compared with 0.3329 and 0.7314, respectively.
- The unrestricted regression matrix yields  $\begin{bmatrix} -0.1705 & 0.3692 & 0.1162 & 0.0740 \\ -0.0313 & 0.2386 & 0.4072 & 0.2386 \end{bmatrix}$ .

Note that the submatrices required to have rank one for identifiability in fact have rank two. In other words, the relationships that the PLS mode A proxies satisfy are at variance with the true model (and these relationships will be different again for mode B and the various modes C). This result implies that different estimation methods will yield, as a rule, different probability limits (for some of the statistical implications, see Dijkstra, 1981, 1983, 2010).

Clearly, when not corrected for inconsistency, PLS tends to give the wrong idea of the relative sizes of the parameters of the underlying covariance structure. Although the signs are correct, this is generally not guaranteed. Counterexamples for regression models are easy to provide (cf. Dijkstra, 1981, pp. 75–76; a program generating counterexamples is available from the first author on request).

## 3.4. Results

### 3.4.1. The unrestricted correlations between the latent variables

This section reports some results specific for PLSc: an estimate of the unrestricted correlation matrix of the latent variables is obtained as an intermediate product. The third and fourth columns of Table 1 contain the average and standard deviation of the estimates, based on 10,000 normal samples of size 300. For the fifth and sixth columns, 500 samples of size 300 were generated, and 1000 bootstrap samples were taken from each.

Evidently, for the model analyzed, PLSc yields estimators for the correlations between the latent variables that are close to unbiased. The bootstrap can also be trusted to produce relatively stable estimates of the standard errors with a slight downward bias.

### 3.4.2. The structural parameters

As previously stated, 2SLS is used for the “heart” of the model—the linear relationships between the latent variables. First, the differences between 2SLS and FIML are reported as applied to the latent variables *as if* they could be directly observed. Next, the analysis of the full model will reveal the size of the price to be paid for indirect observations, for both contenders.

**Table 2**

2SLS and FIML applied to the true latent variable scores.

Parameter	True value	2SLS		FIML	
		Mean	Std	Mean	Std
$\gamma_{11}$	−0.3000	−0.2998	0.0488	−0.2996	0.0487
$\gamma_{12}$	0.5000	0.5001	0.0616	0.5015	0.0606
$\gamma_{23}$	0.5000	0.5004	0.0269	0.5005	0.0271
$\gamma_{24}$	0.2500	0.2500	0.0244	0.2498	0.0242
$\beta_{12}$	0.2500	0.2501	0.0755	0.2480	0.0752
$\beta_{21}$	0.5000	0.5001	0.0423	0.5001	0.0431
Convergence:		100.00%		99.64%	

**Table 3**

Structural model results for 300 observations, multivariate normality (estimates based on 10,000 simulation samples).

Parameter	True value	FIML		PLSc		PLS (with 2SLS)	
		Mean	Std	Mean	Std	Mean	Std
$\gamma_{11}$	−0.3000	−0.3045	0.0920	−0.2990	0.0905	−0.1615	0.0549
$\gamma_{12}$	0.5000	0.5103	0.1263	0.4994	0.1155	0.2994	0.0681
$\gamma_{23}$	0.5000	0.5027	0.0818	0.5002	0.0751	0.3619	0.0509
$\gamma_{24}$	0.2500	0.2496	0.0718	0.2502	0.0732	0.2188	0.0493
$\beta_{12}$	0.2500	0.2452	0.1367	0.2526	0.1315	0.2944	0.1073
$\beta_{21}$	0.5000	0.5040	0.1390	0.4983	0.1323	0.5927	0.1335
Convergence:		99.82%		100.00%		100.00%	

**Table 4**

Structural model results for 300 observations, kurtosis = 1.7124 (estimates based on 10,000 simulation samples).

Parameter	True value	FIML		PLSc	
		Mean	Std	Mean	Std
$\gamma_{11}$	−0.3000	−0.3023	0.1134	−0.2976	0.1133
$\gamma_{12}$	0.5000	0.5097	0.1502	0.4987	0.1472
$\gamma_{23}$	0.5000	0.5018	0.0924	0.4998	0.0949
$\gamma_{24}$	0.2500	0.2492	0.0879	0.2478	0.0926
$\beta_{12}$	0.2500	0.2412	0.1703	0.2550	0.1675
$\beta_{21}$	0.5000	0.5040	0.1780	0.5015	0.1686
Convergence:		98.42%		100.00%	

Table 2 shows the results for 2SLS and FIML, obtained from 10,000 normal samples of size 300 (9964 samples for FIML, attributable to 36 cases of non-convergence). As is shown, the FIML and the 2SLS estimators are virtually unbiased. The results are very similar, and for this model there does not appear to be much to choose from between the two competitors.

What happens if the variables are no longer observable? Specifically, what happens when we have only three modestly correlated indicators per latent variable?

We look at the leading case – multivariate normality – first. Table 3 presents the FIML results for 9982 samples (for 18 samples, FIML did not converge). A slight bias in the estimates is visible, and the price for unobservability is quite substantial: roughly, the standard errors are doubled or tripled. Table 3 also presents the PLSc results. Again, as in the case of the observable variables, a very similar performance of PLSc and FIML is observed. We note that traditional PLS, using 2SLS for the structural model, produces relatively stable estimators, approximately unbiased for its probability limits (which deviate strongly from the true values). Recall that, given the misrepresentation of the correlation structure of the latent variables by traditional PLS, other simultaneous equation estimation methods (and other modes) will yield different central values for the estimators. In the sequel, we focus on PLSc and FIML.

Next, the performance of FIML and PLSc in the case of moderately non-normal data is examined. If the indicators are scaled by the factor  $\sqrt{|Z|} \cdot \sqrt{\frac{\pi}{2}}$ , which means that the kurtosis is increased by a modest 1.7124, both methods suffer, but again in similar ways. Table 4 shows the results for FIML and PLSc. Again, both estimators are of similar quality.

Finally, the performance of FIML and PLSc in the case of highly kurtotic data is examined. Specifically, the indicators are rescaled by the normal variable  $Z$ , increasing the kurtosis to six. For the first time, PLSc showed apparent convergence problems for two out of 10,000 samples; in 96.7% of the cases, PLSc converged in no more than seven steps. FIML struggled a bit more: for 1067 out of 10,000 samples, convergence did not occur. Table 5 presents the results for FIML and PLSc. Again, both techniques exhibit a comparable performance. The FIML estimates appear slightly more biased than PLSc.

**Table 5**

Structural model results for 300 observations, kurtosis = 6 (estimates based on 10,000 simulation samples).

Parameter	True value	FIML		PLSc	
		Mean	Std	Mean	Std
$\gamma_{11}$	−0.3000	−0.3108	0.1612	−0.2976	0.1604
$\gamma_{12}$	0.5000	0.5289	0.2118	0.5018	0.2101
$\gamma_{23}$	0.5000	0.5063	0.1319	0.4970	0.1338
$\gamma_{24}$	0.2500	0.2516	0.1245	0.2501	0.1317
$\beta_{12}$	0.2500	0.2306	0.2449	0.2574	0.2362
$\beta_{21}$	0.5000	0.4949	0.2489	0.4991	0.2422
Convergence:		89.33%		99.98%	

**Table 6**

Average standard errors of loadings per latent variable.

Latent variable	Std of loadings	
	FIML	PLSc
$\eta_1$	0.0426	0.0670
$\eta_2$	0.0415	0.0507
$\eta_3$	0.0405	0.0524
$\eta_4$	0.0420	0.0569
$\eta_5$	0.0409	0.0734
$\eta_6$	0.0365	0.0403

Rescaling the indicator vectors by  $\sqrt{\text{abs}(Z)} \cdot \sqrt{\frac{\pi}{2}}$  leads to standard errors for *both* methods that are approximately  $\sqrt{3\pi/6} \approx 1.25$  times as large as for the normal case. Rescaling by  $Z$  effectively multiplies the normal case standard errors by approximately  $\sqrt{3} \approx 1.73$ . These numbers are no accident; they agree with the (asymptotic) corrections for non-normal kurtosis in covariance structure analysis (cf. Bentler and Dijkstra, 1985). Therefore, the ratios between corresponding standard errors of PLSc and FIML are constant for the analyzed conditions.

It seems fair to say for this setup that both “contenders” tend to produce estimators for the structural parameters of the same or comparable quality. Of course, under non-normality (and especially with non-independence between errors and latent variables), one cannot be sure that the usual way to calculate standard errors (as derived from the information matrix) yields proper estimates. With FIML, one could anticipate underestimation for the non-normal models with the increased kurtosis considered here, as borne out by some experiments (not displayed). One should use one of the known correction methods available in standard software, such as EQS, or simply the bootstrap. The latter is the obvious choice for PLSc.

### 3.4.3. The loadings

Based on 10,000 normal samples (9982 for FIML), the 18 loading estimators are essentially unbiased for both methods. The FIML average values differ from 0.7000 by no more than 0.0010 (and by 0.0005 on average), and for PLSc these numbers are 0.0040 and 0.0020, respectively. The FIML standard errors are definitely smaller. Because they appear to be equal for the loadings on the same latent variable, and the same is true for PLSc, averages of standard errors per latent variable are reported (see Table 6).

These results are reinforced for the non-normal distributions. Unbiasedness is not affected, but the standard errors are. They are not reported because the same phenomenon occurs as for the structural form parameters (and even more clearly). Rescaling the indicator vectors by  $\sqrt{\text{abs}(Z)} \cdot \sqrt{\frac{\pi}{2}}$  leads again to standard errors for *both* methods that are all approximately  $\sqrt{3\pi/6} \approx 1.25$  times as large as for the normal case. Additionally, rescaling by  $Z$  effectively multiplies the normal case standard errors by  $\sqrt{3} \approx 1.73$ . Therefore, the ratios between the corresponding standard errors of PLSc and FIML are also constant for the analyzed conditions.

With three modestly correlated indicators per latent variable, the full information method lives up to its expectations as far as the loadings are concerned. This situation may change with a more unfavorable ratio of observations to indicators (doubling the indicators, for example) but has yet to be pursued. This study did not attempt to experiment with the simple PLSc algorithm using other selections of latent variables in the regressions leading to the sample proxies, or changing the coefficients in those regressions. Additional work is therefore required.

### 3.4.4. Overall test-of-fit

The experiment also provided insights into the performance of the overall test-of-fit. Some bootstrap results are reported for the normal distribution. In the case of non-normal kurtosis values, one may need to “downweight” the observations, as in Yuan and Hayashi (2003), which may require delicate fine-tuning. Because this first study is meant to be illustrative of the possibilities of PLSc, a more elaborate analysis is postponed to another occasion. In addition, we investigate the relative

**Table 7**  
Rejection probabilities of FIML and two discrepancy functions for PLSc.

Observations	Rejection probabilities			
	Nominal	FIML	$d_G$	$d_{LS}$
300	10.0%	13.4%	4.2%	4.9%
	5.0%	7.8%	1.5%	1.8%
	2.5%	4.4%	0.5%	0.5%
600	10.0%	11.5%	7.7%	8.2%
	5.0%	5.8%	3.9%	4.0%
	2.5%	3.3%	1.8%	1.7%
1200	10.0%	10.8%	9.1%	10.3%
	5.0%	5.7%	5.0%	4.3%
	2.5%	2.9%	2.8%	1.9%

frequency with which the true model is rejected when the fit statistic exceeds the bootstrap-based estimate of a conventional quantile, as well as the distribution of the probability values (the relative frequency with which the bootstrapped distances exceed the observed distance). Power analyses are deferred to the future. Yuan and Bentler (1998) and Yuan and Hayashi (2003) should be reviewed for careful discussions of the issues involved in testing whether a covariance model fits.

Two distance functions are considered:

$$d_{LS} := \frac{1}{2} \text{trace} (S - \hat{S})^2, \quad (35)$$

the squared Euclidean distance, and

$$d_G := \frac{1}{2} \sum_{k=1}^{\text{\#indicators}} (\log(\varphi_k))^2, \quad (36)$$

the geodesic distance. Here,  $\varphi_k$  is the  $k$ th eigenvalue of  $S^{-1}\hat{S}$ . Both distances are zero if and only if the model fits perfectly:  $\hat{S} = S$ . They belong to different classes:  $d_{LS}$  cannot be expressed in terms of the eigenvalues. The geodesic distance is one of Swain's (1975) fitting functions; normalized, they are asymptotically equivalent to the likelihood ratio statistic.  $d_G$  is characterized by the property that its minimization with respect to the parameters leaves the generalized variance intact (given scale invariance of the model, see Dijkstra, 1990). When evaluated at PLSc's  $\hat{S}$ , of course it cannot be expected to follow a  $\chi^2$  distribution.

We generated 1000 normal samples of size 300. For each sample, the implied correlation matrix  $\hat{S}$  and its distance to the sample correlation matrix  $S$ ,  $d(\hat{S}, S)$  were calculated, both for least squares and the geodesic distance. The observation vectors were pre-multiplied by  $\hat{S}^{-\frac{1}{2}} S^{-\frac{1}{2}}$  and 1000 bootstrap samples of the transformed values were generated, with the model refitted and distances re-calculated for each bootstrap sample. We note whether the observed distance  $d(\hat{S}, S)$  exceeds certain quantiles of the empirical distribution function of the bootstrapped distances. If it does, a false alarm occurs. Ideally, the average number of false alarms agree with the theoretical values.

The upper third of Table 7 contains the empirical rejection probabilities for several levels of nominal rejection probabilities based on 300 observations. These are not bad (see Yuan and Bentler, 1998; Yuan and Hayashi, 2003), but they are clearly too small. The test is more cautious than desired. Apparently, when the sample does not fit too well, the bootstrapped transformed sample tends to be worse. Also (not shown), the histogram of the probability values is not uniform, and its shape is like a parabola with the maximum value in the middle. It helps to increase the sample size. The same exercise for  $n = 600$  results in the values reported in the middle part of Table 7. Now the histogram of the probability values is definitely closer to uniform. The lower part of Table 7 shows that by doubling again,  $n = 1200$ , the rejection probabilities are almost right. Again, the histogram of probability values is closer to uniform. The balanced bootstrap yielded very similar results.

Following a suggestion by a reviewer, we also looked at the behavior of the normalized Chi-square statistic, based on FIML, compared with the critical value as determined by its limiting distribution (see Table 7). Whereas PLSc is too conservative, we note that FIML displays the well-known tendency to be too skeptical (see Hu et al., 1992; Yuan and Bentler, 1998). Perhaps it is worthwhile to try to construct an appropriate combination of the tests (other than coin tossing, which will be controversial on grounds of substance) in order to protect oneself to errors of the first kind.

### 3.4.5. Summary of experimental results

For the model analyzed in this section, PLSc provides as good a picture of the all-important structural parameters as FIML. In addition, the unrestricted correlations allow proper estimation, and there are grounds to believe that a correct test-of-fit is well within reach. The loadings are more difficult to estimate accurately, which gives FIML a definite advantage. On the whole, one can certainly maintain that the results are encouraging.

**Table 8**

Structural model results for 100 observations, 9 indicators per construct, normal data (estimates based on 10,000 simulation samples).

Parameter	True value	FIML		PLSc	
		Mean	Std	Mean	Std
$\gamma_{11}$	−0.3000	−0.3017	0.1109	−0.2943	0.1131
$\gamma_{12}$	0.5000	0.5130	0.1426	0.5002	0.1394
$\gamma_{23}$	0.5000	0.5039	0.0771	0.4968	0.0775
$\gamma_{24}$	0.2500	0.2520	0.0738	0.2473	0.0745
$\beta_{12}$	0.2500	0.2358	0.1749	0.2660	0.1655
$\beta_{21}$	0.5000	0.4908	0.1389	0.4968	0.1335
Convergence:		96.55%		100.00%	

#### 4. Limitations and avenues for future research

The point of departure in this paper is a class of linear structural equation models, where linearly linked latent variables are measured indirectly by at least two indicators, with each indicator loading on one latent variable. The information between blocks of observables is conveyed solely by the latent variables. In PLS, this setup is known as the basic design (Wold, 1982), and serves as a platform for the development and test of estimation methods. In practice, one may encounter situations that are conceptually less clear-cut, and loading matrices as well as covariance matrices of errors of different latent variables may have structures that are more complicated than the ones handled in this paper. How these situations affect PLSc is still unclear, but it certainly requires analysis. Other topics for further research were alluded to in the text. These topics include power analyses for overall tests-of-fit and tests of robustness against structural misspecification to be compared with full information methods. The power analyses and robustness tests are related, and they are best studied in conjunction, allowing a proper assessment of the expected trade-off between them. An investigation into the effects of skewness and kurtosis on convergence, speed, and stability of the algorithm will also be interesting.

Existing simulation studies on the performance of PLS with respect to reflective measurement should be replicated using consistent PLS. Many characteristics that were already explored for PLS, such as statistical power (Chin and Newsted, 1999; Goodhue et al., 2006), parameter accuracy and convergence behavior (Reinartz et al., 2009; Henseler, 2010), or performance relative to other component-based structural equation modeling techniques (Hwang et al., 2010; Lu et al., 2011; McDonald, 1996; Tenenhaus, 2008), need to be investigated for consistent PLS.

For PLSc, 2SLS was suggested for the relationships between the latent variables, but there are many alternatives. A class of alternatives that suggests itself is the set of minimum distance or GLS estimators based on unrestricted CAN estimators for the regression matrix. They will probably be computationally more expensive and less robust, but perhaps more efficient (asymptotically) under correct specification.

For a long time, analysts have criticized PLS' lack of a global goodness-of-fit measure. The approaches suggested by PLS researchers (GoF and relative GoF, see Tenenhaus et al., 2004, 2005; Esposito Vinzi et al., 2010) assess the average correlations of the estimated relationships between the proxies and between the latter and the indicators. In contrast, our test-of-fit is in the spirit of classical covariance structure analysis, where a generalized distance between the sample covariance matrix and a structured, theoretical covariance matrix is assessed in order to measure the appropriateness of the structural assumptions.

A reviewer offered the conjecture that PLSc may have an advantage over covariance-based SEM for small samples and complex models. We tend to agree. Programs like LISREL, or FIML generally, try to maximize the fit of highly nonlinear models (in terms of restrictions on moments or distribution functions) to sample data, allowing every structural constraint to have an impact. The outcomes of FIML may therefore be more sensitive to small sample variation than those of PLSc. The latter first constructs proxies without the constraints of the model relationships between the latent variables, and then, as implemented here, estimates each equation separately. Some complementary work by Dijkstra and Schermelleh-Engel (2013) on nonlinear structural models, with products and squares of latent variables, lends support to the conjecture: the maximum likelihood method (LMS) effectively broke down when the number of indicators became large. PLSc had no trouble producing structural parameter estimates, and they were in fact more accurate (in terms of standard deviations) compared with the estimators based on a smaller number of indicators.

Clearly, a serious investigation into the extent to which the conjecture is valid will be a major project. We have offered a relatively simple study, as a prelude to more extensive research. Consider the model as before, but now *triple* the number of indicators per latent variable (nine instead of three), with the same loadings (0.7), and with the same structural parameters, and reduce the sample size by two-thirds to one hundred. So we generated 10,000 samples of size 100 from a 54-dimensional normal population. For PLSc it took less than one and a half minute to generate 10,000 samples and estimate the parameters on each of them. PLSc converged for all samples. In 97% of the cases it needed 4, 5 or 6 iterations (mostly 5, in 85% of the samples), and all correction factors were positive. FIML was clearly challenged, it took 25 h to do the entire exercise. The estimation results are however very similar (see Table 8): The structural parameter estimators are again essentially unbiased, and their standard errors are definitely smaller than before, when we used only three indicators per latent variable.

At the very least, the numerical expediency of PLSc makes it a natural candidate for computationally expensive analyses, such as (double) bootstraps (for example for overall tests, or confidence intervals based on bootstrap bias corrected estimators) and bootstrap-jackknife combinations. Beyond the demands of practice, and despite the open research challenges as listed above, we would maintain that PLSc warrants serious consideration in structural equation estimation.

## References

- Bekker, P.A., Dijkstra, T.K., 1990. On the nature and number of the constraints on the reduced form as implied by the structural form. *Econometrica* 58 (2), 507–514.
- Bekker, P.A., Merckens, A., Wansbeek, T.J., 1994. Identification, Equivalent Models, and Computer Algebra. Academic Press, Boston.
- Bentler, P.M., Dijkstra, T.K., 1985. Efficient estimation via linearization in structural models. In: Krishnaiah, P.R. (Ed.), *Multivariate Analysis VI*. North-Holland, Amsterdam, pp. 9–42.
- Boardman, A., Hui, B., Wold, H., 1981. The partial least squares-fix point method of estimating interdependent systems with latent variables. *Comm. Statist. Theory Methods* 10 (7), 613–639.
- Bry, X., Redont, P., Verron, T., Cazes, P., 2012. THEME-SEER: a multidimensional exploratory technique to analyze a structural model using an extended covariance criterion. *J. Chemom.* 26 (5), 158–169.
- Chin, W.W., 1998. Issues and opinion on structural equation modeling. *MIS Q.* 22 (1), vii–xvi.
- Chin, W.W., Dibbern, J., 2010. An introduction to a permutation based procedure for multi-group PLS analysis: results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA. In: Vinzi, V.E., Chin, W.W., Henseler, J., Wang, H. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications*. In: *Computational Statistics*, vol. II. Springer, Heidelberg, Dordrecht, London, New York, pp. 171–193.
- Chin, W.W., Marcolin, B.L., Newsted, P.R., 2003. A partial least squares latent variable modeling approach for measuring interaction effects. Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Inf. Syst. Res.* 14 (2), 189–217.
- Chin, W.W., Newsted, P.R., 1999. Structural equation modeling analysis with small samples using partial least squares. In: Hoyle, R.H. (Ed.), *Statistical Strategies for Small Sample Research*. Sage Publications, Thousand Oaks, London, New Delhi, pp. 307–341 (Chapter 12).
- Chin, W.W., Thatcher, J.B., Wright, R.T., 2014. Assessing common method bias: problems with the ULMC technique. *MIS Q.* (forthcoming).
- Dijkstra, T.K., 1981. Latent variables in linear stochastic models: Reflections on “maximum likelihood” and “partial least squares” methods (Ph.D. thesis), Groningen University, Groningen, a second edition was published in 1985 by Sociometric Research Foundation.
- Dijkstra, T.K., 1983. Some comments on maximum likelihood and partial least squares methods. *J. Econometrics* 22 (1–2), 67–90.
- Dijkstra, T.K., 1990. Some properties of estimated scale invariant covariance structures. *Psychometrika* 55 (2), 327–336.
- Dijkstra, T.K., 2010. Latent variables and indices: Herman Wold’s basic design and partial least squares. In: Vinzi, V.E., Chin, W.W., Henseler, J., Wang, H. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications*. In: *Computational Statistics*, vol. II. Springer, Heidelberg, Dordrecht, London, New York, pp. 23–46.
- Dijkstra, T.K., 2013. A note on how to make PLS consistent. Working Paper. <http://www.rug.nl/staff/t.k.dijkstra/how-to-make-pls-consistent.pdf>.
- Dijkstra, T.K., Schermelleh-Engel, K., 2013. Consistent partial least squares for nonlinear structural equation models. *Psychometrika* <http://dx.doi.org/10.1007/s11336-013-9370-0>.
- Esposito Vinzi, V., Trinchera, L., Amato, S., 2010. PLS path modeling: from foundations to recent developments and open issues for model assessment and improvement. In: Vinzi, V.E., Chin, W.W., Henseler, J., Wang, H. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications*. In: *Computational Statistics*, vol. II. Springer, Heidelberg, Dordrecht, London, New York, pp. 47–82.
- Fleishman, A., 1978. A method for simulating non-normal distributions. *Psychometrika* 43 (4), 521–532.
- Fornell, C., Bookstein, F.L., 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *J. Mark. Res.* 19 (4), 440–452.
- Gefen, D., Rigdon, E.E., Straub, D., 2011. An update and extension to SEM guidelines for administrative and social science research. *MIS Q.* 35 (2), iii–xiv.
- Goodhue, D., Lewis, W., Thompson, R., 2006. PLS, small sample size, and statistical power in MIS research. In: *HICSS’06. Proceedings of the 39th Annual Hawaii International Conference on System Sciences*. IEEE, p. 202b.
- Haavelmo, T., 1944. The probability approach in econometrics. *Econometrica* 12 (Supplement), iii–vi, 1–115.
- Hair, J.F., Sarstedt, M., Ringle, C.M., Mena, J.A., 2012. An assessment of the use of partial least squares structural equation modeling in marketing research. *J. Acad. Mark. Sci.* 40 (3), 414–433.
- Henseler, J., 2010. On the convergence of the partial least squares path modeling algorithm. *Comput. Statist.* 25 (1), 107–120.
- Henseler, J., 2012. Why generalized structured component analysis is not universally preferable to structural equation modeling. *J. Acad. Mark. Sci.* 40 (3), 402–413.
- Henseler, J., Chin, W.W., 2010. A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Struct. Equ. Model.* 17 (1), 82–109.
- Henseler, J., Fassott, G., Dijkstra, T.K., Wilson, B., 2012. Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. *Eur. J. Inf. Syst.* 21 (1), 99–112.
- Henseler, J., Ringle, C.M., Sinkovics, R.R., 2009. The use of partial least squares path modeling in international marketing. *Adv. Int. Mark.* 20, 277–319.
- Henseler, J., Sarstedt, M., 2013. Goodness-of-fit indices for partial least squares path modeling. *Comput. Statist.* 28 (2), 565–580.
- Hsieh, J.J., Rai, A., Keil, M., 2008. Understanding digital inequality: comparing continued use behavioral models of the socio-economically advantaged and disadvantaged. *MIS Q.* 32 (1), 97–126.
- Hu, L.-t., Bentler, P.M., Kano, Y., 1992. Can test statistics in covariance structure analysis be trusted? *Psychol. Bull.* 112 (2), 351–362.
- Hulland, J., 1999. Use of partial least squares (PLS) in strategic management research. A review of four recent studies. *Strateg. Manage. J.* 20, 195–204.
- Hwang, H., Malhotra, N., Kim, Y., Tomiuk, M., Hong, S., 2010. A comparative study on parameter recovery of three approaches to structural equation modeling. *J. Mark. Res.* 47 (4), 699–712.
- Krijnen, W.P., Dijkstra, T.K., Gill, R.D., 1998. Conditions for factor (in) determinacy in factor analysis. *Psychometrika* 63 (4), 359–367.
- Liang, H., Saraf, N., Hu, Q., Xue, Y., 2007. Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. *MIS Q.* 31 (1), 59–87.
- Lohmöller, J.B., 1989. *Latent Variable Path Modeling with Partial Least Squares*. Physica, Heidelberg.
- Lu, L., Kwan, E., Thomas, D., Cedzynski, M., 2011. Two new methods for estimating structural equation models: an illustration and a comparison with two established methods. *Int. J. Res. Mark.* 28 (3), 258–268.
- McDonald, R., 1996. Path analysis with composite variables. *Multivariate Behav. Res.* 31 (2), 239–270.
- Money, K., Hillenbrand, C., Henseler, J., da Camara, N., 2012. Exploring unanticipated consequences of strategy amongst stakeholder segments: the case of a European revenue service. *Long Range Plann.* 45 (5–6), 395–423.
- Pearl, J., 2009. *Causality: Models, Reasoning, and Inference*, second ed. Cambridge University Press.
- Peng, D.X., Lai, F., 2012. Using partial least squares in operations management research: a practical guideline and summary of past research. *J. Oper. Manage.* 30 (6), 467–480.
- Reinartz, W., Haenlein, M., Henseler, J., 2009. An empirical comparison of the efficacy of covariance-based and variance-based SEM. *Int. J. Res. Mark.* 26 (4), 332–344.
- Ringle, C.M., Sarstedt, M., Straub, D.W., 2012. A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Q.* 36 (1), iii–xiv.
- Sarstedt, M., Henseler, J., Ringle, C.M., 2011. Multigroup analysis in partial least squares (PLS) path modeling: alternative methods and empirical results. *Adv. Int. Mark.* 22, 195–218.



- Summers, R., 1965. A capital intensive approach to the small sample properties of various simultaneous equation estimators. *Econometrica* 33 (1), 1–41.
- Swain, A., 1975. A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika* 40 (3), 315–335.
- Tenenhaus, M., 2008. Component-based structural equation modelling. *Total Qual. Manage. Bus. Excell.* 19 (7–8), 871–886.
- Tenenhaus, M., Amato, S., Esposito Vinzi, V., 2004. A global goodness-of-fit index for PLS structural equation modelling. In: *Proceedings of the XLII SIS Scientific Meeting*, pp. 739–742.
- Tenenhaus, A., Tenenhaus, M., 2011. Regularized generalized canonical correlation analysis. *Psychometrika* 76 (2), 257–284.
- Tenenhaus, M., Vinzi, V.E., Chatelin, Y.M., Lauro, C., 2005. PLS path modeling. *Comput. Statist. Data Anal.* 48 (1), 159–205.
- Vale, C., Maurelli, V., 1983. Simulating multivariate nonnormal distributions. *Psychometrika* 48 (3), 465–471.
- Vinzi, V.E., Chin, W.W., Henseler, J., Wang, H. (Eds.), 2010. *Handbook of Partial Least Squares: Concepts, Methods, and Applications*. In: *Computational Statistics*, vol. II. Springer, Heidelberg, Dordrecht, London, New York.
- Wetzels, M., Odekerken-Schröder, G., van Oppen, C., 2009. Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration. *MIS Q.* 33 (1), 177–195.
- Wold, H.O.A., 1966. Non-linear estimation by iterative least squares procedures. In: David, F.N. (Ed.), *Research Papers in Statistics*. Wiley, London, New York, Sydney, pp. 411–444.
- Wold, H.O.A., 1975. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. In: Gani, J. (Ed.), *Perspectives in Probability and Statistics. Papers in Honour of M. S. Bartlett on the Occasion of his Sixty-Fifth Birthday*. Applied Probability Trust, Sheffield, pp. 117–142.
- Wold, H.O.A., 1982. Soft modelling: the basic design and some extensions. In: Jöreskog, K.G., Wold, H.O.A. (Eds.), *Systems Under Indirect Observation. Causality, Structure, Prediction*. Vol. II. North-Holland, Amsterdam, New York, Oxford, pp. 1–54.
- Yuan, K., Bentler, P., 1998. Normal theory based test statistics in structural equation modelling. *British J. Math. Statist. Psych.* 51 (2), 289–309.
- Yuan, K., Hayashi, K., 2003. Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *British J. Math. Statist. Psych.* 56 (1), 93–110.